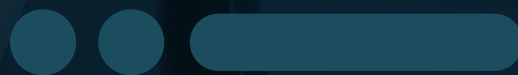
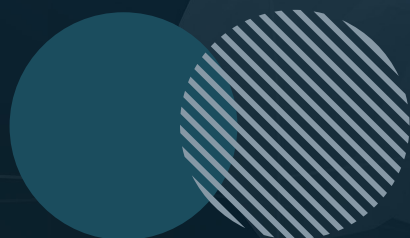




CEAG E MS

GUIA PARA AVALIAÇÃO DE POLÍTICAS PÚBLICAS

ANÁLISE QUANTITATIVA DE IMPACTO
DE POLÍTICAS – PROADI-SUS



Projeto: Estruturação de Monitoramento na gestão estratégica de projetos e cooperações do Ministério da Saúde (MS)

Centro de Estudos Avançados de Governo e Administração Pública
Universidade de Brasília

2026

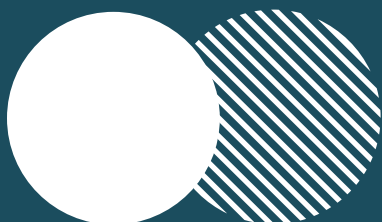


CEAG E MS



GUIA PARA AVALIAÇÃO DE POLÍTICAS PÚBLICAS

ANÁLISE QUANTITATIVA DE IMPACTO
DE POLÍTICAS – PROADI-SUS



Projeto: Estruturação de Monitoramento na gestão estratégica de projetos e cooperações do Ministério da Saúde (MS)

Centro de Estudos Avançados de Governo e Administração Pública
Universidade de Brasília

2026





PROJETO DE PESQUISA

Estruturação de Monitoramento na gestão estratégica de projetos e cooperações do Ministério da Saúde (MS)

EQUIPE ACADÊMICA E TÉCNICA – UNIVERSIDADE DE BRASÍLIA (UNB)

Prof. Dr. Luiz Guilherme de Oliveira

Prof.a Dra. Doriana Daroit

Prof.a Dra. Susan Elizabeth Martins Cesar de Oliveira

Prof.a Dra. Gabriela Borges Antunes

Prof.a Dra. Fátima de Souza Freire

Prof. Dr. Arnaldo Mauerberg Junior

Prof. Dr. Antônio Nascimento Junior

Prof. Dr. Leonardo Cavalcanti da Silva

Prof. Dr. Paulo Carlos Du Pin Calmon

Prof. Dr. Roberto Góes Ellery Júnior

Prof. Dr. Victor Gomes e Silva

Prof. Dr. Wladimir Ganzelevitch Gramacho

ADMINISTRATIVO

Ms. Simone Braga Farias

Me. Wilson Alves Borba Junior

BOLSISTAS DO PROGRAMA DE DOUTORADO

Paula Daniella Prado Ramos

Solana Irene Loch Zandonai

Fabiana Bandeira dos Santos





APRESENTAÇÃO

O presente guia contempla os registros referentes ao desenvolvimento da Meta 2 que se concentra em propostas para implementação de modelo, métodos e técnicas para monitoramento e avaliação de programas e projetos de cooperação técnica do Ministério da Saúde (MS), especificamente no Produto 2.3: Proposta de métodos e técnicas de avaliação para cada objetivo dos programas e projetos de cooperação técnica do MS.

O objetivo principal desse guia é apoiar técnicos do MS na avaliação de políticas públicas, oferecendo um conjunto de técnicas para estimar o impacto das ações implementadas no âmbito dos programas Proadi-SUS, Pronas/PCD e Pronon. A avaliação de impacto é uma ferramenta essencial para medir os efeitos de uma política sobre a população e entender sua eficácia, auxiliando na tomada de decisões baseadas em evidências.





SUMÁRIO

1. INTRODUÇÃO 5

2. ETAPAS PRELIMINARES 7

- 2.1 Elaborando as questões 7
- 2.2 Construção dos indicadores 9

3. MÉTODOS DE AVALIAÇÃO 11

- 3.1 Experimento Controlado 13
- 3.2 Métodos Não-Experimentais 14
 - 3.2.1 Estimador de Diferenças em Diferenças (DiD) 15
 - 3.2.2 Estimador de Variáveis Instrumentais 17
 - 3.2.3 Regressões com descontinuidade 18
 - 3.2.4 Matching 19
 - 3.2.5 Regressões múltiplas 21
 - 3.2.6 Controle Sintético 22
- 3.3 Outros Métodos 23
 - 3.3.1 Comparação “antes e depois” 24
 - 3.3.2 Diferença simples 25

4. APRESENTAÇÃO DOS RESULTADOS DA AVALIAÇÃO 27

5. CONSIDERAÇÕES FINAIS 29





1. INTRODUÇÃO

O objetivo desse documento é oferecer um conjunto de técnicas que possam ser utilizadas para avaliar políticas públicas, especialmente no âmbito do Ministério da Saúde. O público-alvo são técnicos do Ministério da Saúde que estejam envolvidos com avaliação de impacto das políticas executadas pelo ministério. Embora alguns aspectos do planejamento de uma avaliação de impacto sejam bastante técnicos, este documento busca fornecer ao leitor um arcabouço geral para participar com sucesso de uma ampla gama de avaliações de impacto.

As avaliações de impacto estimam o efeito de uma política sobre um conjunto de resultados de interesse. Exemplos incluem estimar o impacto de uma campanha de vacinação em massa na redução da incidência de doenças como sarampo ou poliomielite em uma determinada população, avaliar o impacto de uma campanha de conscientização sobre higiene (como lavagem das mãos) na diminuição de doenças diarreicas em comunidades vulneráveis, e medir o efeito de um programa de distribuição gratuita de medicamentos para hipertensão ou diabetes na melhora da qualidade de vida e na redução de complicações associadas. O objetivo final é ser capaz de avaliar as mudanças no(s) resultado(s) de interesse que são atribuíveis ao programa em avaliação,

para que possamos determinar pelo que o programa é responsável e quão eficaz ele foi em alcançar seus objetivos. Dito de outra forma, o problema de avaliar o impacto de uma política é o problema da inferência causal.

A tarefa fundamental da inferência causal é determinar se uma relação de causa e efeito existe entre uma intervenção (como um programa ou tratamento) e um resultado observado. Isso envolve distinguir os efeitos diretamente atribuíveis à intervenção daqueles que poderiam ter ocorrido devido a outros fatores externos ou condições preexistentes. Em essência, busca-se responder à pergunta: “O que teria acontecido com os indivíduos ou a população se a intervenção não tivesse sido implementada?” Esse cenário contrafactual, porém, é impossível de observar diretamente, o que torna a inferência causal um desafio analítico complexo.

Para superar essa dificuldade, os métodos de inferência causal, como experimentos controlados randomizados ou técnicas estatísticas avançadas (como pareamento ou variáveis instrumentais), são utilizados para estimar o efeito causal. Esses métodos tentam criar uma comparação válida entre o grupo que recebeu a intervenção e um grupo equivalente que não a recebeu, isolando assim o impacto real da intervenção. O objetivo é garantir que as conclusões sobre a eficácia de um programa ou política sejam baseadas em evidências confiáveis, permitindo decisões informadas sobre sua continuidade, ajuste ou expansão.

No restante desse documento serão apresentadas abordagens para avaliação de impacto utilizadas nas avaliações de políticas públicas no Brasil e no mundo. A apresentação não tem como objetivo esgotar as questões técnicas envolvendo cada abordagem, para isso existe vasta literatura disponível. Em vez disso, o objetivo é focar nas decisões-chave que precisam ser tomadas durante a etapa de planejamento da avaliação de impacto, e em particular nos papéis que tanto o responsável por encomendar a avaliação quanto o especialista técnico encarregado de conduzi-la devem desempenhar.





2. ETAPAS PRELIMINARES

2.1 ELABORANDO AS QUESTÕES

A elaboração de questões relevantes para uma avaliação de impacto de políticas públicas utilizando inferência causal é um passo crucial para garantir que os resultados sejam úteis e informativos. O governo, como principal responsável pela formulação e implementação das políticas, desempenha um papel central ao definir os objetivos gerais da avaliação. Ele deve identificar claramente o que deseja compreender, por exemplo, se uma política de transferência de renda reduz a pobreza ou se um programa de saúde melhora indicadores de bem-estar. Essas questões precisam ser específicas, mensuráveis e alinhadas aos propósitos da política, refletindo as prioridades do poder público e as necessidades da sociedade.

Os especialistas encarregados de elaborar a avaliação, como pesquisadores ou analistas de dados, têm a responsabilidade de traduzir os objetivos do governo em perguntas tecnicamente viáveis para a inferência causal. Isso envolve determinar quais resultados de interesse (como taxas de escolaridade ou mortalidade) serão medidos e como o

efeito causal da política será isolado de outros fatores. Eles devem propor questões que permitam estimar o contrafactual, o que teria ocorrido na ausência da política, e sugerir métodos apropriados, como ensaios randomizados ou análises quasi-experimentais. O diálogo entre especialistas e governo é essencial para garantir que as questões sejam realistas, considerando limitações de dados e recursos disponíveis.

O público-alvo da política, ou seja, os indivíduos ou comunidades diretamente afetadas por ela, também deve ser considerado na formulação das questões. Suas experiências e perspectivas ajudam a identificar aspectos da política que podem não ser imediatamente evidentes para o governo ou os especialistas. Por exemplo, uma questão relevante pode surgir ao perguntar se os beneficiários percebem mudanças em sua qualidade de vida ou acesso a serviços, o que pode direcionar a avaliação para resultados práticos e não apenas teóricos. Envolver o público-alvo, por meio de consultas ou pesquisas, assegura que as questões reflitam impactos reais e não apenas metas abstratas.

Além disso, as questões devem ser elaboradas de forma a equilibrar os interesses do governo, a expertise técnica dos especialistas e as necessidades do público-alvo. Por exemplo, enquanto o governo pode querer saber se uma política é custo-efetiva, os especialistas podem priorizar a robustez metodológica, e o público-alvo pode se preocupar com a acessibilidade da política. Uma questão bem elaborada, como “Qual é o impacto de um programa de capacitação profissional na empregabilidade de jovens em áreas urbanas vulneráveis?”, consegue atender a essas diferentes demandas, fornecendo respostas úteis para todos os envolvidos.

Por fim, a construção dessas questões é um processo iterativo que exige colaboração contínua. O governo deve estar disposto a ajustar suas expectativas com base nas limitações apontadas pelos especialistas, enquanto estes precisam adaptar seus métodos às prioridades políticas e às vozes do público-alvo. O sucesso de uma avaliação de impacto depende de questões que não apenas permitam inferências causais sólidas, mas também gerem evidências práticas para aprimorar políticas públicas, beneficiando tanto os tomadores de decisão quanto a população atendida. Assim, o alinhamento entre esses três atores, governo, especialistas e público, é o fundamento para uma avaliação relevante e eficaz.



2.2 CONSTRUÇÃO DOS INDICADORES

A construção de indicadores-chave de resultado para a avaliação de políticas públicas é um processo essencial para medir o sucesso e o impacto dessas iniciativas. Esses indicadores devem ser cuidadosamente elaborados para refletir os objetivos da política e permitir uma análise clara e objetiva. O governo, como principal formulador da política, precisa definir quais mudanças deseja observar, por exemplo, a redução da taxa de desemprego ou o aumento da cobertura vacinal. Para isso, os indicadores devem seguir características específicas, como o modelo SMART¹: serem específicos (focados em um aspecto claro), mensuráveis (quantificáveis), alcançáveis (viáveis), relevantes (alinhados aos objetivos) e temporais (limitados a um período definido).

Uma característica essencial dos indicadores é sua consistência com a teoria que sustenta a política pública. Isso significa que eles devem estar fundamentados em uma lógica causal explícita, conhecida como “teoria da mudança”, que explica como a intervenção deve gerar os resultados esperados. Por exemplo, se uma política de treinamento profissional visa aumentar a empregabilidade, o indicador-chave, como a porcentagem de participantes empregados após seis meses, deve refletir essa relação causal. Especialistas em avaliação, como estatísticos, economistas ou cientistas sociais, desempenham um papel crucial ao garantir que os indicadores sejam teoricamente sólidos e tecnicamente robustos, evitando métricas vagas ou desconexas do propósito da política.

Outro aspecto importante é a necessidade de limitar o número de indicadores. Embora seja tentador medir muitos resultados para capturar todos os efeitos possíveis, um conjunto reduzido e bem selecionado facilita a coleta de dados, a análise e a comunicação dos resultados. O governo e os especialistas devem colaborar para priorizar indicadores que capturem os impactos mais significativos, evitando sobrecarga de informação. Por exemplo, em um programa de saúde materno-infantil, pode-se optar por apenas dois indicadores principais: a taxa de mortalidade infantil e a porcentagem de partos assistidos por profissionais, em vez de dezenas de métricas secundárias.

¹Do inglês: *Specific, Measurable, Achievable, Relevant and Time-bound*.

O público-alvo da política também influencia a construção dos indicadores, pois suas necessidades e realidades ajudam a definir o que é mais relevante. Um indicador SMART deve ser sensível às mudanças que afetam diretamente os beneficiários, por exemplo, em uma política de habitação, medir o número de famílias realocadas para moradias seguras é mais significativo para o público do que apenas o total de recursos gastos. Envolver representantes do público-alvo no processo, por meio de consultas ou grupos focais, pode garantir que os indicadores reflitam benefícios tangíveis e não apenas metas administrativas.

Por fim, a construção de indicadores-chave exige um equilíbrio entre ambição e praticidade. O governo pode buscar resultados amplos, mas os especialistas devem garantir que os indicadores sejam viáveis com os dados disponíveis e os prazos estabelecidos. Revisões periódicas são recomendadas para ajustar os indicadores conforme a política evolui ou novos desafios surgem. Quando indicadores bem elaborados, que respeitem as características SMART, são consistentes com a teoria e limitados em número, tornam-se ferramentas poderosas para avaliar o impacto de políticas públicas, oferecendo clareza para os tomadores de decisão e benefícios reais para a sociedade.





3. MÉTODOS DE AVALIAÇÃO

O conceito de contrafactual é fundamental na inferência causal, pois representa o cenário hipotético que descreve o que teria acontecido com um indivíduo ou grupo caso uma intervenção específica não tivesse sido aplicada. Em termos simples, é a resposta à pergunta: “Qual seria o resultado se a política ou programa não tivesse ocorrido?” Essa ideia é crucial porque, na realidade, só podemos observar o que aconteceu após a intervenção (o fato), mas não o que teria ocorrido sem ela (o contrafactual). Estabelecer esse cenário alternativo permite estimar o verdadeiro efeito causal de uma ação, isolando-o de outros fatores.

Na prática, como o contrafactual não pode ser diretamente observado, os pesquisadores utilizam métodos para aproximá-lo. Em experimentos controlados randomizados, por exemplo, o grupo de controle serve como uma proxy do contrafactual, pois é composto por indivíduos que não recebem a intervenção, mas são comparáveis ao grupo de tratamento em termos de características iniciais. Já em estudos observacionais, técnicas como pareamento (matching) ou modelos de regressão são empregadas para construir um contrafactual estimado, ajustando variáveis que poderiam influenciar o resultado. A qualidade dessa aproximação é o que determina a validade da inferência causal.

Um desafio central ao trabalhar com o contrafactual é garantir que ele seja realista e represente fielmente as condições que existiriam na ausência da intervenção. Isso exige que os pesquisadores controlem variáveis confundidoras, fatores externos que poderiam afetar o resultado independentemente da política ou programa avaliado. Por exemplo, ao avaliar o impacto de um programa de saúde na redução de doenças, mudanças econômicas ou sazonais podem influenciar os resultados. Se o contrafactual não levar esses fatores em conta, o efeito atribuído à intervenção pode ser superestimado ou subestimado, comprometendo a análise.

Em políticas públicas, o uso do contrafactual é especialmente valioso para justificar investimentos e ajustes em programas. Ao comparar os resultados observados com o cenário contrafactual estimado, os tomadores de decisão podem entender o que a política realmente mudou — por exemplo, quantos empregos foram criados por um programa de capacitação que não teriam surgido sem ele. Apesar de sua importância, a construção de um contrafactual confiável exige dados robustos, métodos rigorosos e, muitas vezes, suposições teóricas claras, o que torna a inferência causal uma tarefa complexa, mas indispensável para avaliações eficazes.

A utilização de experimentos controlados aleatórios na avaliação de políticas públicas enfrenta significativas dificuldades políticas e morais, que muitas vezes limitam sua aplicação. Politicamente, a randomização pode gerar resistência de gestores e comunidades, pois implica oferecer uma intervenção a apenas parte da população-alvo, enquanto outros são deliberadamente excluídos para formar o grupo de controle, o que pode ser percebido como injusto ou discriminatório. Moralmente, essa exclusão levanta dilemas éticos, especialmente quando a política envolve bens essenciais, como saúde, educação ou alimentação. Negar acesso a um grupo em nome da ciência pode ser considerado inaceitável, sobretudo se os benefícios da intervenção já são presumidos. Esses conflitos frequentemente levam à pressão por soluções alternativas, como implementações graduais ou estudos observacionais, que, embora menos rigorosos para inferência causal, atendem melhor às demandas de equidade e aceitação pública.



3.1 EXPERIMENTO CONTROLADO

O método de experimento controlado, frequentemente chamado de ensaio controlado randomizado (ECR), é uma abordagem amplamente utilizada para inferência causal. Consiste em dividir uma população em dois grupos: o grupo de tratamento, que recebe a intervenção (como um programa ou política), e o grupo de controle, que não a recebe. A alocação dos indivíduos a esses grupos é feita de forma aleatória, garantindo que as diferenças observadas nos resultados entre os grupos possam ser atribuídas à intervenção, e não a fatores pré-existentes. Esse método busca replicar o “contrafactual” usando o grupo de controle como referência.

As principais hipóteses estatísticas do ECR incluem a aleatoriedade da atribuição, que assegura que os grupos sejam comparáveis em características observáveis e não observáveis antes da intervenção. Outra hipótese é a independência entre os grupos, ou seja, o tratamento de um grupo não deve afetar o outro (conhecido como ausência de “efeitos de *spillover*”). Além disso, assume-se que a intervenção é a única diferença sistemática entre os grupos, permitindo que as diferenças nos resultados, como taxas de escolaridade ou saúde, sejam interpretadas como o efeito causal da política. Testes estatísticos, como o teste t ou regressões, são usados para verificar se essas diferenças são significativas.

Na prática, a implementação de ECRs em ciências sociais enfrenta diversas dificuldades. Uma delas é a questão ética: em políticas públicas, pode ser moralmente problemático negar a intervenção a um grupo de controle, especialmente se ela envolve benefícios essenciais, como acesso a saúde ou educação. Por exemplo, avaliar um programa de alimentação escolar excluindo algumas crianças pode gerar críticas e resistência. Isso leva à busca por alternativas, como listas de espera ou implementação gradual, mas essas soluções nem sempre mantêm a pureza do desenho experimental.

Outra dificuldade é a contaminação entre grupos, especialmente em contextos sociais onde indivíduos interagem. Se membros do grupo de controle são influenciados pelo grupo de tratamento, por exemplo, ao compartilhar informações ou recursos, o efeito isolado da intervenção fica comprometido. Em políticas públicas, como campanhas de conscientização, esse “efeito de *spillover*” é difícil de evitar, pois as men-

sagens podem se espalhar além do grupo-alvo. Isso exige ajustes no desenho do estudo ou o uso de unidades maiores, como comunidades inteiras, o que aumenta os custos e a complexidade.

A representatividade da amostra também é um desafio. Para que os resultados de um ECR sejam generalizáveis, a população estudada deve refletir as características do público-alvo da política em larga escala. No entanto, experimentos em pequena escala, como em uma única cidade, podem não capturar variações regionais ou culturais relevantes. Além disso, a implementação real de políticas públicas muitas vezes diverge do ambiente controlado de um experimento, introduzindo variáveis externas, como mudanças econômicas ou políticas, que afetam os resultados e dificultam a inferência causal.

Por fim, os custos e o tempo necessários para conduzir um ECR são obstáculos significativos. Em ciências sociais e políticas públicas, onde os efeitos podem levar anos para se manifestar (como em programas educacionais), os experimentos exigem recursos substanciais e paciência dos tomadores de decisão. Apesar dessas dificuldades, o ECR permanece um padrão-ouro para inferência causal quando bem implementado, pois oferece evidências robustas sobre a eficácia de intervenções, ajudando governos e especialistas a tomar decisões baseadas em dados sólidos, ainda que com adaptações contextuais.

3.2 MÉTODOS NÃO-EXPERIMENTAIS

Os métodos não experimentais para inferência causal são amplamente utilizados quando experimentos controlados randomizados não são viáveis, seja por questões éticas, logísticas ou financeiras, especialmente na avaliação de políticas públicas. Esses métodos, baseados em dados observacionais, incluem técnicas como o pareamento (*matching*), variáveis instrumentais, diferenças-em-diferenças (*difference-in-differences*) e regressão descontínua (*regression discontinuity*). O objetivo é estimar o contrafactual controlando variáveis confundidoras que poderiam influenciar o resultado. Por exemplo, ao avaliar o impacto de um aumento do salário-mínimo, o método de diferenças-em-diferenças compara a evolução do emprego em regiões afetadas e não afetadas antes e depois da política, isolando seu efeito.



Embora esses métodos ofereçam flexibilidade, sua principal dificuldade reside na dependência de suposições fortes e na qualidade dos dados disponíveis. No pareamento, por exemplo, é necessário assumir que todas as variáveis relevantes foram observadas e equilibradas entre os grupos comparados, o que nem sempre é realista, pois fatores não observáveis podem gerar vieses. Da mesma forma, o uso de variáveis instrumentais exige encontrar um instrumento válido, uma variável que afeta a intervenção, mas não o resultado diretamente, o que pode ser tecnicamente desafiador. Apesar dessas limitações, os métodos não experimentais são valiosos para analisar políticas já implementadas ou em contextos em que a randomização é impraticável, fornecendo evidências causais quando combinados com análises robustas e interpretações cuidadosas.

3.2.1 Estimador de Diferenças em Diferenças (DiD)

O estimador de diferenças em diferenças (*difference-in-differences*, ou DiD) é um método não experimental amplamente utilizado para inferência causal na avaliação de políticas públicas. Ele compara a evolução de um resultado de interesse ao longo do tempo entre dois grupos: um grupo afetado pela intervenção (tratamento) e outro que não foi (controle). A lógica do método é subtrair a diferença nos resultados antes e depois da intervenção no grupo de controle da diferença observada no grupo de tratamento, isolando assim o efeito causal da política. Por exemplo, ao avaliar o impacto de uma nova lei trabalhista sobre o emprego, o DiD analisa como as taxas de emprego mudam nas regiões onde a lei foi implementada em comparação com regiões onde não houve mudança, controlando tendências temporais comuns.

Uma das principais hipóteses do DiD é a “tendência paralela” (*parallel trends assumption*), que presume que, na ausência da intervenção, os grupos de tratamento e controle teriam seguido trajetórias semelhantes no resultado de interesse. Essa suposição é crucial, pois permite atribuir qualquer divergência pós-intervenção à política avaliada. Para verificar essa hipótese, os pesquisadores frequentemente analisam os dados pré-intervenção, examinando se as tendências nos dois grupos eram consistentes antes da mudança. Se essa condição não for atendida, talvez devido a diferenças econômicas preexistentes entre os

grupos, o estimador pode gerar conclusões enviesadas, superestimando ou subestimando o efeito da política.

O DiD é particularmente útil em políticas públicas porque aproveita dados longitudinais já disponíveis, como registros administrativos ou pesquisas populacionais, eliminando a necessidade de experimentos caros ou randomização. Uma aplicação comum é na avaliação de reformas educacionais: suponha que um estado introduza um programa de reforço escolar. O DiD pode comparar as notas dos alunos nesse estado com as de outro estado sem o programa, antes e depois da implementação, para estimar o impacto no desempenho escolar. Essa abordagem é flexível e pode ser ajustada com regressões para incluir variáveis de controle, como renda ou tamanho da escola, aumentando a precisão da análise.

Apesar de suas vantagens, o DiD enfrenta desafios práticos e teóricos. Um problema frequente é a presença de choques externos que afetam os grupos de maneira desigual após a intervenção, como crises econômicas ou mudanças legislativas paralelas, que podem confundir os resultados. Além disso, a escolha do grupo de controle é crítica: ele deve ser suficientemente semelhante ao grupo de tratamento em características relevantes, mas não afetadas pela política. Em contextos complexos, como políticas nacionais com múltiplos fatores em jogo, esses requisitos podem ser difíceis de satisfazer, exigindo testes de robustez e validação cuidadosa dos resultados.

Em resumo, o estimador de diferenças em diferenças é uma ferramenta poderosa para avaliar políticas públicas, oferecendo uma alternativa viável aos métodos experimentais quando a randomização não é possível. Suas aplicações vão desde a análise de impactos de subsídios agrícolas até a avaliação de programas de saúde, como a introdução de campanhas de vacinação. Com dados adequados e suposições bem fundamentadas, o DiD permite aos tomadores de decisão entenderem os efeitos reais de suas ações, contribuindo para o desenho de políticas mais eficazes e baseadas em evidências, ainda que demande atenção rigorosa às suas limitações metodológicas.



3.2.2 Estimador de Variáveis Instrumentais

O estimador de variáveis instrumentais (VI) é uma técnica estatística utilizada para lidar com o problema de endogeneidade em modelos de regressão, que ocorre quando uma variável explicativa está correlacionada com o termo de erro. Essa correlação pode surgir devido a omissão de variáveis relevantes, erro de medição ou simultaneidade entre as variáveis explicativas e a variável dependente. O método de VI resolve esse problema ao utilizar uma variável instrumental, que deve estar correlacionada com a variável explicativa endógena, mas não com o erro do modelo, permitindo a obtenção de estimadores consistentes e não viesados.

Na avaliação de políticas públicas, a endogeneidade é um desafio comum, pois muitas políticas não são implementadas de forma aleatória. Por exemplo, a alocação de recursos educacionais pode depender do nível socioeconômico da população, tornando difícil identificar o efeito causal do investimento na educação sobre os resultados dos alunos. O uso de VI pode ajudar a superar esse problema ao encontrar um fator externo que afete a variável explicativa sem estar diretamente relacionado ao resultado analisado. Um exemplo clássico é o uso da distância de uma escola como variável instrumental para medir o impacto da educação sobre os salários.

Outro exemplo de aplicação do estimador de VI em políticas públicas está na área da saúde. Suponha que um pesquisador queira estimar o efeito de um programa de vacinação sobre a taxa de mortalidade infantil. Se a adesão ao programa for influenciada por fatores individuais não observáveis, como preocupações com a saúde, a estimativa direta pode estar enviesada. Uma solução seria utilizar a distância até a unidade de saúde mais próxima como variável instrumental, assumindo que ela afeta a probabilidade de vacinação, mas não influencia diretamente a mortalidade infantil.

Apesar de sua utilidade, a escolha de uma boa variável instrumental é um dos principais desafios do método. Se a variável instrumental for fraca, ou seja, pouco correlacionada com a variável explicativa endógena, os resultados podem ser imprecisos. Além disso, a validade do instrumento precisa ser cuidadosamente testada, pois se houver alguma correlação com o termo de erro, os resultados podem ser viesados.

dos. Dessa forma, o estimador de VI é uma ferramenta poderosa na avaliação de políticas públicas, mas seu uso requer uma escolha criteriosa de instrumentos e verificações estatísticas rigorosas.

3.2.3 Regressões com descontinuidade

A metodologia de Regressão com Descontinuidade (*Regression Discontinuity Design* – RDD) é uma técnica econométrica amplamente utilizada para estimar efeitos causais em situações em que há um critério de corte que determina a participação em um programa ou política pública. O princípio básico do RDD é que indivíduos muito próximos desse ponto de corte são comparáveis, permitindo que se avalie o impacto da intervenção como se fosse um experimento quase-aleatório. Isso ocorre porque aqueles logo acima e logo abaixo do limite enfrentam condições similares, exceto pela exposição ao tratamento, possibilitando uma inferência causal mais robusta.

O RDD pode ser aplicado em diversas áreas da avaliação de políticas públicas. Por exemplo, em programas educacionais, políticas que concedem bolsas de estudo ou acesso a escolas de qualidade com base em uma nota mínima em um exame são candidatas ideais para essa abordagem. Comparando o desempenho de estudantes que ficaram logo acima e logo abaixo do corte, é possível estimar o efeito da bolsa ou do acesso diferenciado na aprendizagem, empregabilidade e outros resultados de interesse. Como esses alunos tendem a ser muito semelhantes em características observáveis e não observáveis, a descontinuidade no desfecho pode ser atribuída ao programa em questão.

Outro campo de aplicação do RDD é na avaliação de programas sociais e de transferência de renda. Suponha que um benefício seja concedido apenas a famílias com renda abaixo de um determinado valor. Ao comparar indicadores socioeconômicos das famílias que ficaram ligeiramente abaixo e ligeiramente acima desse limite, pode-se medir o efeito do programa na redução da pobreza, no consumo ou no bem-estar dos beneficiários. Esse método é particularmente útil porque evita viés de seleção associado a decisões individuais de participação no programa.

Na área da saúde, a RDD tem sido usada para avaliar o impacto de políticas que estabelecem critérios objetivos para acesso a determinados tratamentos ou serviços médicos. Um exemplo é a elegibilidade



para determinado medicamento gratuito baseada na idade ou em um nível específico de gravidade da doença. Se os pacientes logo acima e logo abaixo desse limiar forem comparáveis, a diferença nos seus desfechos de saúde pode ser atribuída ao acesso ao tratamento, permitindo uma avaliação mais precisa da eficácia da política pública.

Apesar das vantagens, a metodologia de RDD também apresenta desafios. Um dos principais é garantir que os indivíduos não manipulem o critério de corte para se beneficiar da política, o que poderia comprometer a validade do experimento quase-natural. Além disso, a generalização dos resultados pode ser limitada, pois a estimativa obtida se aplica apenas aos indivíduos próximos ao ponto de corte e pode não ser representativa de toda a população elegível para a política.

Em conclusão, as regressões com descontinuidade são uma ferramenta poderosa para avaliação de políticas públicas, fornecendo estimativas causais robustas em contextos em que há um critério de elegibilidade bem definido. No entanto, sua aplicação exige atenção a possíveis manipulações do ponto de corte e à validade externa dos resultados. Quando bem implementada, essa abordagem permite que formuladores de políticas tomem decisões mais informadas sobre a efetividade de programas e intervenções governamentais.

3.2.4 Matching

O método de *matching* é uma técnica estatística amplamente utilizada na avaliação de políticas públicas para estimar efeitos causais quando não há um experimento aleatorizado. O objetivo principal desse método é comparar indivíduos que participaram de uma determinada política pública (grupo tratado) com indivíduos não participantes (grupo de controle), garantindo que ambos sejam o mais semelhantes possível em relação a características observáveis. Dessa forma, o *matching* busca reduzir o viés de seleção, permitindo uma comparação mais justa entre os grupos e uma melhor estimativa do impacto da política analisada.

Uma das principais vantagens do *matching* é sua capacidade de construir um grupo de controle mais apropriado, evitando comparações inadequadas entre indivíduos com características muito distintas. Ao parear indivíduos com base em variáveis como idade, nível de esco-

laridade, renda ou localização, o método assegura que diferenças nos resultados não sejam atribuídas a fatores externos, mas sim ao efeito da política pública. Além disso, ao contrário de métodos experimentais, o *matching* pode ser aplicado em bases de dados observacionais, tornando-o uma alternativa viável quando experimentos aleatorizados não são possíveis ou éticos.

Outro ponto forte do método é a flexibilidade na escolha da técnica de pareamento. Existem diversas abordagens de *matching*, como *propensity score matching* (PSM), que pareia indivíduos com base na probabilidade de participação no programa, e *nearest neighbor matching*, que encontra indivíduos com características mais próximas dentro do grupo de controle. Essa diversidade de técnicas permite que pesquisadores escolham a abordagem mais adequada para seu contexto específico, melhorando a validade da análise.

Apesar de suas vantagens, o método de *matching* também apresenta limitações importantes. Uma das principais fragilidades é que ele só controla por características observáveis, ou seja, fatores não mensurados ou não disponíveis nos dados podem continuar gerando viés na estimativa do impacto. Isso significa que, se houver variáveis não observadas que influenciam tanto a participação na política quanto os resultados analisados, a estimativa ainda pode ser enviesada. Além disso, o *matching* pode ser sensível à especificação do modelo e à escolha das variáveis utilizadas para pareamento, exigindo um cuidado rigoroso na seleção de dados.

Por fim, outro desafio do *matching* é a necessidade de encontrar um número suficiente de indivíduos comparáveis no grupo de controle. Em alguns casos, pode ser difícil encontrar pares adequados para todos os participantes do programa, especialmente quando a política pública atende a um grupo muito específico. Isso pode reduzir a precisão dos resultados ou até mesmo impossibilitar a aplicação do método. Apesar dessas limitações, quando utilizado corretamente e combinado com outras técnicas, o *matching* é uma ferramenta poderosa para avaliação de políticas públicas, fornecendo estimativas mais confiáveis e auxiliando na formulação de intervenções mais eficazes.



3.2.5 Regressões múltiplas

As regressões múltiplas são amplamente utilizadas na avaliação de políticas públicas como uma ferramenta estatística para estimar relações entre variáveis e medir o impacto de programas governamentais. Esse método permite analisar o efeito de uma variável explicativa sobre um resultado de interesse enquanto controla por outras variáveis que podem influenciar a relação. Por exemplo, ao avaliar o impacto de um programa de transferência de renda sobre o desempenho escolar, é possível incluir variáveis como idade, escolaridade dos pais e localização geográfica para isolar melhor o efeito da política em questão.

Uma das principais vantagens das regressões múltiplas é a sua flexibilidade e aplicabilidade a uma ampla gama de contextos e dados observacionais. Como o método permite controlar por diversas variáveis simultaneamente, ele ajuda a reduzir problemas de variáveis omitidas que poderiam enviesar os resultados. Além disso, as regressões múltiplas podem ser usadas com diferentes tipos de dados, incluindo séries temporais, dados em painel e cortes transversais, tornando-se uma ferramenta versátil na análise de políticas públicas.

Outra grande força do método é a facilidade de interpretação e a disponibilidade de técnicas para testar a robustez dos resultados. Estatísticas como o R^2 , testes de significância e intervalos de confiança ajudam a avaliar a qualidade do modelo e a precisão das estimativas. Além disso, a inclusão de interações entre variáveis permite investigar se o impacto da política varia entre diferentes grupos da população, fornecendo informações mais detalhadas para a formulação de políticas mais direcionadas.

Apesar de suas vantagens, as regressões múltiplas apresentam limitações significativas. Uma das principais fraquezas é a dificuldade de estabelecer relações de causa e efeito apenas com base em correlações. Se uma variável explicativa estiver correlacionada com o erro do modelo, o estimador pode estar enviesado devido a problemas de endogeneidade. Isso pode ocorrer devido à omissão de variáveis relevantes, simultaneidade entre variáveis ou erro de medição, o que compromete a validade das conclusões.

Outro desafio das regressões múltiplas é a possibilidade de multicolinearidade, que ocorre quando duas ou mais variáveis explicativas

estão altamente correlacionadas entre si. Isso pode tornar difícil distinguir os efeitos individuais de cada variável e levar a coeficientes instáveis e estatisticamente insignificantes. Além disso, a escolha inadequada de variáveis pode gerar modelos mal especificados, resultando em inferências equivocadas sobre o impacto da política pública analisada.

Por fim, a qualidade das estimativas obtidas em regressões múltiplas depende fortemente da qualidade dos dados disponíveis. Se os dados forem incompletos, imprecisos ou não representativos, os resultados podem ser distorcidos e levar a conclusões erradas. Portanto, embora as regressões múltiplas sejam uma ferramenta poderosa para a avaliação de políticas públicas, elas devem ser aplicadas com rigor metodológico, preferencialmente combinadas com outras abordagens para garantir maior validade e robustez na identificação de efeitos causais.

3.2.6 Controle Sintético

O método de controle sintético é uma técnica econométrica utilizada na avaliação de políticas públicas para estimar o impacto de uma intervenção quando não há um grupo de controle ideal disponível. Ele funciona criando uma combinação ponderada de unidades não tratadas (como estados, municípios ou países) para formar um “controle sintético” que simula o comportamento da unidade tratada caso a política não tivesse sido implementada. Essa abordagem tem sido amplamente utilizada para avaliar políticas como leis ambientais, reformas educacionais e programas de incentivo econômico.

Uma das principais vantagens do controle sintético é sua capacidade de construir um grupo de comparação mais apropriado do que métodos tradicionais, como diferenças simples ou diferenças-em-diferenças, especialmente quando a política pública afeta uma única unidade ou poucas unidades. Em vez de comparar diretamente a unidade tratada com um único grupo de controle, o método utiliza uma combinação de múltiplas unidades ponderadas para criar um contrafactual mais robusto. Isso reduz o viés causado por diferenças estruturais entre o grupo tratado e o controle.

Outra força do método é a transparência e a replicabilidade da construção do grupo sintético. Como os pesos atribuídos às unidades de controle são determinados de maneira formal e baseada em caracte-



terísticas observáveis antes da política ser implementada, há menor risco de manipulação subjetiva na seleção do grupo de comparação. Além disso, o método permite a visualização clara da evolução das tendências antes e depois da intervenção, facilitando a comunicação dos resultados para formuladores de políticas e a sociedade.

No entanto, o controle sintético também apresenta limitações. Um dos principais desafios é a necessidade de um conjunto amplo de unidades não tratadas para construir um controle adequado. Se houver poucas unidades disponíveis para compor o grupo sintético, o modelo pode não conseguir criar uma boa aproximação do contrafactual, tornando a estimativa pouco confiável. Além disso, o método é sensível à escolha das variáveis utilizadas para construir a combinação ponderada, podendo gerar estimativas instáveis caso variáveis importantes sejam omitidas.

Do ponto de vista estatístico, um dos principais problemas do controle sintético é a inferência. Como geralmente há apenas uma unidade tratada e um grupo sintético construído para comparação, as técnicas estatísticas tradicionais para estimar a incerteza dos resultados (como testes de significância) nem sempre se aplicam diretamente. Para contornar esse problema, muitas avaliações utilizam placebos ou testes de permutação para verificar a robustez dos resultados, mas a interpretação dos intervalos de confiança ainda pode ser desafiadora. Apesar dessas limitações, o controle sintético é uma ferramenta poderosa para avaliar políticas públicas em cenários onde outras abordagens não são viáveis, fornecendo estimativas mais confiáveis do impacto de intervenções governamentais.

3.3 OUTROS MÉTODOS

Algumas vezes não é possível usar os métodos descritos nessa seção. Isso decorre de fatores como limitações na coleta de dados, falta de condições para implementar um ensaio controlado randomizado, ou mesmo falta de tempo hábil para aplicar técnicas estatísticas apropriadas. Nesses casos alguns autores sugerem usar métodos que, apesar de limitados do ponto de vista estatístico, podem dar alguma indicação a respeito dos impactos das políticas analisadas.



3.3.1 Comparação “antes e depois”

O método de comparações antes e depois é uma abordagem simples e amplamente utilizada na avaliação de políticas públicas. Ele consiste em comparar os resultados observados antes da implementação de uma política com os resultados obtidos após sua execução, buscando identificar possíveis mudanças atribuídas à intervenção. Por exemplo, ao avaliar o impacto de um programa de capacitação profissional sobre o emprego, um pesquisador pode comparar a taxa de emprego dos participantes antes e depois do curso para verificar se houve melhora nos índices de empregabilidade.

Uma das principais vantagens desse método é sua facilidade de aplicação. Como ele não exige um grupo de controle, pode ser utilizado em diversas situações em que não há dados disponíveis para comparações com indivíduos ou localidades não afetadas pela política. Além disso, os dados necessários para essa análise geralmente estão disponíveis em registros administrativos ou pesquisas periódicas, reduzindo custos e tempo para a realização da avaliação.

Outra vantagem do método é sua utilidade na análise de impactos imediatos e no monitoramento da evolução de indicadores ao longo do tempo. Ele pode ser útil para gestores públicos que precisam de informações rápidas sobre os primeiros efeitos de uma política e para ajustes em sua implementação. Além disso, se houver dados coletados ao longo de vários períodos, pode-se observar tendências e verificar se a política teve um impacto duradouro.

Apesar de sua simplicidade, essa abordagem apresenta sérias limitações metodológicas. O principal problema é a dificuldade em estabelecer uma relação de causa e efeito entre a política pública e as mudanças observadas nos indicadores. Como muitos outros fatores podem influenciar os resultados ao longo do tempo, não há garantia de que a diferença observada antes e depois seja realmente causada pela intervenção e não por outros eventos simultâneos.

Um problema estatístico fundamental dessa abordagem é a falta de um grupo de controle, o que impede que se isole o efeito real da política. Se a economia estiver crescendo e a taxa de desemprego diminuir naturalmente, por exemplo, um programa de capacitação pode parecer eficaz mesmo que seu impacto real tenha sido pequeno. Esse



problema é conhecido como viés de fatores externos ou viés de tempo, pois eventos externos podem influenciar os resultados independentemente da política pública.

Outra questão estatística relevante é a reversão à média, que ocorre quando resultados extremos tendem a se aproximar da média ao longo do tempo, independentemente da intervenção. Isso pode levar a interpretações equivocadas sobre a eficácia de uma política. Por exemplo, se um programa de segurança pública for implementado após um pico de criminalidade, uma eventual queda nos crimes pode ocorrer naturalmente e não devido à política.

Por fim, o método de comparações antes e depois pode levar a conclusões equivocadas se houver mudanças no perfil da população ou na forma como os dados são coletados. Se os critérios de medição dos indicadores forem alterados ao longo do tempo, a comparação pode ser distorcida. Assim, embora seja um método acessível e útil para análises preliminares, ele deve ser usado com cautela e, sempre que possível, complementado por outras metodologias mais robustas, como o uso de grupos de controle ou métodos econométricos que controlem por fatores externos.

3.3.2 Diferença simples

O método de diferenças simples é uma abordagem utilizada na avaliação de políticas públicas que compara os resultados entre um grupo que recebeu a intervenção e um grupo que não recebeu. Essa comparação permite estimar o impacto da política pública analisando se há uma diferença nos indicadores entre os dois grupos após a implementação da política. Por exemplo, para avaliar um programa de incentivo à educação, pode-se comparar a taxa de conclusão escolar entre alunos que receberam e não receberam o benefício.

Uma das principais vantagens do método de diferenças simples é sua simplicidade e facilidade de interpretação. Ele permite obter uma estimativa direta do efeito da política sem a necessidade de dados históricos ou modelos econométricos sofisticados. Além disso, quando aplicado corretamente, pode oferecer uma primeira indicação do impacto da intervenção, sendo útil para gestores públicos que precisam de respostas rápidas sobre a efetividade de programas governamentais.

No entanto, esse método apresenta sérias limitações, especialmente no que diz respeito ao viés de seleção. Se os grupos tratados e não tratados forem sistematicamente diferentes antes da implementação da política, a estimativa obtida pode estar enviesada. Por exemplo, se apenas escolas de alta qualidade participarem de um programa de financiamento educacional, qualquer melhora no desempenho dos alunos pode ser mais resultado das características dessas escolas do que do financiamento em si. A falta de controle por essas diferenças pré-existentes pode levar a conclusões equivocadas sobre a efetividade da política.

Do ponto de vista estatístico, um dos principais problemas da abordagem de diferenças simples é a impossibilidade de separar corretamente os efeitos causais da política de outros fatores que podem influenciar os resultados. Variáveis não observadas, como diferenças culturais ou socioeconômicas entre os grupos comparados, podem afetar os resultados e gerar um viés de confundimento. Dessa forma, para que o método seja mais confiável, é essencial garantir que os grupos comparados sejam suficientemente semelhantes, o que pode ser feito com técnicas mais sofisticadas, como pareamento (*matching*) ou a abordagem de diferenças-em-diferenças, que controla melhor esses problemas.





4. APRESENTAÇÃO DOS RESULTADOS DA AVALIAÇÃO

Ao reportar uma avaliação de política pública baseada em inferência causal, é fundamental garantir que os resultados sejam apresentados de forma clara, precisa e transparente. Isso significa explicar a metodologia utilizada, justificar a escolha do método de inferência causal (como variáveis instrumentais, diferenças-em-diferenças ou regressão com descontinuidade) e discutir as principais limitações do estudo. O relatório deve incluir informações detalhadas sobre os dados utilizados, as fontes de informação e os critérios para a seleção da amostra, garantindo que os leitores possam compreender a robustez da análise.

Uma consideração essencial ao comunicar os resultados é o cuidado com a interpretação das conclusões, especialmente no contexto político. Políticas públicas muitas vezes estão associadas a interesses governamentais e institucionais, e uma avaliação que sugira impactos negativos pode gerar resistência por parte de gestores e tomadores de decisão. Para evitar distorções, é importante apresentar os achados de maneira neutra e baseada em evidências, destacando tanto os efeitos positivos quanto os desafios ou impactos limitados da política. A transparência na comunicação dos resultados aumenta a credibilidade da avaliação e facilita o uso das evidências para a formulação de políticas mais eficazes.

Questões éticas também são fundamentais na divulgação de uma avaliação de política pública. Se a pesquisa envolveu dados sensíveis ou informações pessoais dos participantes, é essencial garantir a confidencialidade e a privacidade dos indivíduos analisados. Além disso, a interpretação dos resultados deve ser feita com responsabilidade, evitando conclusões precipitadas ou afirmações exageradas que possam levar a decisões equivocadas. Recomenda-se a adoção de diretrizes éticas reconhecidas, como aquelas estabelecidas por comitês de ética em pesquisa e organismos internacionais que regulam estudos com impacto social.

Outro ponto crucial na comunicação dos resultados é a consideração dos efeitos sobre o público-alvo da política. Uma avaliação que evidencia impactos negativos pode gerar preocupações e resistência entre os beneficiários, especialmente se houver risco de cortes ou modificações na política em questão. Assim, ao reportar os achados, é importante contextualizar os resultados e, quando possível, sugerir ajustes ou melhorias na política, em vez de apenas apontar falhas. Esse cuidado pode contribuir para um debate mais construtivo sobre a efetividade da intervenção, facilitando sua aceitação e aprimoramento.

Além disso, ao apresentar os resultados para diferentes audiências é recomendável adaptar a linguagem e os formatos de comunicação. Para gestores públicos, um sumário executivo com os principais achados e recomendações práticas pode ser mais eficaz do que um relatório técnico detalhado. Já para o público acadêmico, um relatório mais aprofundado, com metodologia e análise estatística detalhada, pode ser necessário. A clareza na comunicação ajuda a garantir que os resultados sejam compreendidos e utilizados de forma apropriada.

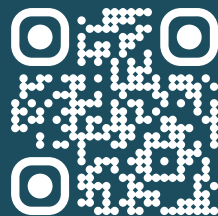
Por fim, a avaliação de políticas públicas deve sempre estar alinhada ao objetivo de contribuir para a melhoria das ações governamentais e para o bem-estar da população. Relatar os resultados com rigor metodológico, responsabilidade ética e sensibilidade política permite que a inferência causal seja utilizada de maneira produtiva, promovendo o aprimoramento das políticas públicas e a tomada de decisões baseadas em evidências.





5. CONSIDERAÇÕES FINAIS

Esse guia teve como objetivo apresentar uma introdução geral às técnicas de avaliação de políticas públicas. Foi feita uma escolha deliberada por não utilizar de modelos matemáticos ou estatísticos para priorizar a intuição bem como as limitações de cada técnica. No relatório final esse guia aparecerá como uma etapa introdutória do método de controle sintético que foi apresentado na oficina realizada na UnB.



ceag.unb.br



[@unb_oficial](https://twitter.com/unb_oficial)



[@ceag_unb](https://www.instagram.com/ceag_unb)



ceag@unb.br

